

Analysis of Five Y-Specific Microsatellite Loci in Asian and Pacific Populations

ESTEBAN PARRA,¹ MARK D. SHRIVER,¹ AGUSTINUS SOEMANTRI,²
STEPHEN T. MCGARVEY,³ JOACHIM HUNDRIESER,⁴
NILMANI SAHA,⁵ AND RANJAN DEKA^{6*}

¹Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania

²Department of Pediatrics, Diponegoro University, Semarang, Indonesia

³Department of Medicine and International Health Institute, Brown University, Providence, Rhode Island

⁴Medizinische Hochschule Hannover, Hannover, Germany

⁵Department of Human Genetics, University of Pittsburgh, Pennsylvania

⁶Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio

KEY WORDS Y-specific microsatellites; genetic polymorphism; Asian and Pacific populations

ABSTRACT We have analyzed five Y-specific microsatellite loci (DYS388, DYS390, DYS391, DYS394, DYS395) in 17 Asian and Pacific populations representing a broad geographical area and different linguistic families, with an emphasis on populations from mainland and insular Southeast Asia. Analysis of gene diversity indicates that several of the studied populations have experienced substantial genetic isolation, and a reduction in male effective sizes (viz. the Northeast Indian populations Nishi, Adi and the Taiwanese aboriginals). The average values of the F_{ST} and (S_T) statistics indicate a high degree of genetic differentiation among these populations at the five Y-specific markers ($F_{ST} = 0.21$ and $(S_T = 0.33$, based on individual loci; $F_{ST} = 0.09$ and $(S_T = 0.36$, based on haplotypes), which conform to the expectation of a fourfold smaller effective size of the Y-linked loci compared with the autosomal loci. Dendrogram and principal coordinates analysis, with few exceptions, show a major separation between mainland and insular populations. Among the mainland populations, the Tibeto-Burman speakers from Northeast India cluster in a well-defined group, supported by high bootstrap values. The Southern Chinese, Northern Thai, So, and Cambodian also are integral to this cluster. The other major cluster is rather heterogeneous and includes, among others, the Austronesian-speaking populations. The Samoans of the Pacific, with a distinctive pattern of allelic distributions, stand as an outlier in the tree and PC representations. Although trends of genetic affinities among ethnically and geographically related populations are evident from the Y-specific microsatellite data, microsatellites are not optimal for deciphering complex migratory patterns of human populations, which could possibly be clarified by using additional and more stable genetic markers. *Am J Phys Anthropol* 110: 1–16, 1999. © 1999 Wiley-Liss, Inc.

The male-specific portion of the human Y chromosome is similar to the mitochondrial genome in that it is haploid and does not recombine during meiosis. However, unlike mitochondrial DNA (mtDNA), the nonrecom

Grant sponsor: National Science Foundation; Grant number: SBR-9600910; Grant sponsor: National Institutes of Health; Grant number: GM-45861.

*Correspondence to: Ranjan Deka, Department of Environmental Health, University of Cincinnati, 3223 Eden Avenue, Cincinnati, Ohio 45267-0056 USA. E-mail: ranjan.deka@uc.edu

Received 29 June 1998; accepted 23 May 1999.

binning region of the Y chromosome (henceforth Y-specific) is paternally transmitted. It thus establishes a paternal lineage, and therefore, Y chromosome polymorphisms can be used to trace the evolutionary past of human populations, albeit via male lines. However, compared to the explosive growth of data that has been amassed for mtDNA and autosomal DNA variation in human populations, data on Y chromosome markers has been far more limited. The Y chromosome has a smaller effective size compared to autosomes, resulting in reduced genetic diversity, and therefore, fewer polymorphic markers have been identified (Jakubiczka et al. 1989; Malaspina et al. 1990; Dorit et al. 1995). In the past few years, however, due largely to advancements in molecular technology, several new polymorphisms have been identified, resulting in an enrichment of Y-specific markers. Recently, a number of studies have demonstrated the potential of the Y chromosome in the interpretation of human evolutionary history (Deka et al. 1996; Hammer 1995; Jobling and Tyler-Smith 1995; Roewer et al., 1996; Santos et al. 1996; Spurdle and Jenkins, 1992; Underhill et al. 1996; Whitefield et al. 1995).

In the present study, we have analyzed genetic polymorphisms at five Y-specific microsatellites (DYS388, DYS390, DYS391, DYS394, DYS395) in 17 Asian and Pacific populations. The data for seven of these populations (Kachari, Japanese, Chinese, Malay, Javanese, Samoan and Papua New Guinea Highlander) were described in a previously published paper in which we had analyzed a sample of 15 world-wide populations (Deka et al., 1996). We present here data from ten additional populations (Nishi, Apatani, Adi, northern Thai, northeastern Thai, Cambodian, So, Orang-Asli, Batak, and Atayal). Although these populations are distributed over a wide geographical area, the emphasis is on the populations from mainland and insular southeast Asia, who belong to a number of diverse linguistic families. Southeast Asia is an interesting area from the anthropological point of view, given that the two major prehistoric migrations that populated the south Pacific had their origins here. More than 40,000 years before present, a first migration brought

hunter-gatherers to Australia and New Guinea (Bellwood, 1978 and 1989), and much later in time, around 6,000 years BP, a second migration began that spread Austronesian languages as far away in the Pacific as Hawaii, New Zealand, and Easter Island (Bellwood, 1989). Although in recent years there have been important advances in our knowledge on genetic relationships among the south Pacific populations (Chen et al., 1992; Sykes et al., 1995; Melton et al., 1995; Lum et al., 1998), information concerning indigenous southeast Asian populations remains limited and incomplete. The main objectives of this study are to contribute to a better understanding of the genetic variation in this area, providing new data on genetic variation at Y-specific microsatellites, and discussing the utility of these markers in evolutionary studies of southeast Asian populations.

SUBJECTS AND METHODS

Population samples

A list of the populations analyzed in the present study is given in Table 1, in which their geographic locations and linguistic affiliations are also indicated. The four Tibeto-Burman speaking populations are from the northeast Indian states of Assam (Kachari) and Arunachal Pradesh (Nishi, Apatani and Adi). The Japanese sample was drawn from the Osaka area of Japan. The Chinese sample was drawn from individuals who migrated from mainland China to Taiwan and are of Han origin. The northern Thai sample was drawn from Chiangmai in northern Thailand, who belong to the Lanna Thai group. The northeastern Thai sample was drawn from the Khon Kaen and the Ubon Ratchathani provinces of northeastern Thailand, and belongs to the Laotian subgroup of Thai languages. The So are an Austroasiatic group and are distributed over the provinces of Sakon Nakhon and Nakhon Phanom of Thailand. The Orang Asli, also known as Semai Senoi, is an aboriginal Austroasiatic speaking group from the peninsular center of Malaysia. Five Austronesian speaking groups are examined, of which Malay, Javanese, and Batak belong to the western Malayo-Polynesian group; the Taiwanese population, Atayal is a member of the Atayalic

TABLE 1. List of the populations analyzed in the present study, with their geographical location and linguistic affiliation

Population	Geographical location	Linguistic affiliation ¹
Kachari	Northeast India	Sino-Tibetan (Tibeto-Burman)
Nishi	Northeast India	Sino-Tibetan (Tibeto-Burman)
Apatani	Northeast India	Sino-Tibetan (Tibeto-Burman)
Adi	Northeast India	Sino-Tibetan (Tibeto-Burman)
Chinese	South China	Sino-Tibetan (Sinitic)
Japanese	Japan	Japanese
Northern Thai	North Thailand	Daic, Lanna Thai
Northeastern Thai	Northeast Thailand	Daic, Laotian Thai
Cambodian	Cambodia	Austroasiatic
So	Northeast Thailand	Austroasiatic
Orang Asli	Malaysia	Austroasiatic
Malay	Malaysia	Austronesian (Western Malayo-Polynesian)
Javanese	Java	Austronesian (Western Malayo-Polynesian)
Batak	Sumatra	Austronesian (Western Malayo-Polynesian)
Atayal	Taiwan	Austronesian (Atayalic)
Samoa	Samoa	Austronesian (Central-Eastern Malayo-Polynesian)
Highlander	Papua New Guinea	Papuan

¹ Data from Cavalli-Sforza et al. (1994) and Voegelin and Voegelin (1977).

subgroup; and the Samoan sample, who belong to the central eastern Malayo-Polynesian group, was drawn from villages distributed over the islands of western Samoa. The Papuan speaking New Guinea Highlander samples were collected from the northern fringes of Papua New Guinea's central highlands.

DNA analysis

Locus and primer descriptions of the five Y-specific microsatellites (DYS388, DYS390, DYS391, DYS394, and DYS395) are given in Jobling and Tyler-Smith (1995). Polymerase chain reaction (PCR) analysis was performed by amplification of 50 ng of genomic DNA in a 25- μ l reaction mixture containing standard PCR buffer, 200 μ M each dNTP, and 1 unit of *Taq* polymerase. The forward primer was end-labeled using [γ -³³P]ATP and T4 polynucleotide kinase. The amplified products were separated on 6% denaturing polyacrylamide gels. Allelic fragments were visualized by autoradiography, and fragment sizes were determined by comparison with an M13 sequence ladder.

Statistical analysis

Allele frequencies were obtained by simple gene counting. Unbiased estimates of gene diversities (D) and their sampling variances were estimated following Nei (1987), using the Arlequin 1.0 computer package (Schneider et al., 1997). This statistic is

equivalent to heterozygosity for autosomal markers.

D_{SW} and D_A genetic distances were estimated by means of the D_{SW} program, kindly provided by Tatsuya Ohta. D_{SW} has been specifically designed for microsatellite loci, taking into account the stepwise mutation model (Shriver et al., 1995). The D_A distance (Nei et al., 1983), a modification of Cavalli-Sforza and Edwards' (1967) chord distance, has been shown to be highly efficient in obtaining the correct tree topology for microsatellite frequency data (Takezaki and Nei, 1996). Neighbor-joining trees (Saitou and Nei, 1987) were constructed on the basis of both distances, and the degree of support of the branches was tested by a bootstrap analysis based on 1,000 replications. A principal coordinates analysis based on the D_{SW} and D_A distances was performed using NTSYS-pc version 1.70 computer package (Rohlf, 1992).

The degree of population differentiation was analyzed by means of the conventional F-statistics and the ϕ -statistics. The latter incorporates the molecular distances among the microsatellite alleles in the analysis, calculated as the squared difference in the number of repeats. Both analyses were carried out by means of the Arlequin 1.0 computer package (Schneider et al., 1997).

Statistical analysis was carried out both using the frequencies of the individual loci considered independently and using the hap-

lotypes constructed on the basis of the five Y-specific microsatellite loci.

RESULTS

Table 2 shows the allele frequencies and gene diversities at the five Y-specific microsatellite loci in 17 Asian populations, where the most frequent allele at each locus is marked in **boldface**. Our sample sizes range between 7 and 28 individuals, and consequently, the standard errors associated with the estimates of the allele frequencies and gene diversities are quite high. The marker DYS390 shows the highest average gene diversity ($D = 0.65$), and the lowest values correspond to DYS388 and DYS391 ($D = 0.44$ and $D = 0.42$, respectively). The average diversity for most of the populations is higher than 0.50, with the exception of four of the tribal populations included in the analysis, three Tibeto-Burman populations of the Indian subcontinent (Nishi, $D = 0.20$; Adi, $D = 0.29$; Apatani, $D = 0.46$), and the Taiwanese aborigine (Atayal, $D = 0.35$). The gene diversities observed for Nishi, Adi and Atayal are significantly lower than those observed in most of the studied populations (Fig. 1).

The extent of genetic differentiation among these Asian and Pacific populations was determined by means of an analysis of molecular variance (AMOVA), both with and without incorporating the microsatellite mutational model (ϕ -statistics and F-statistics, respectively). The conventional F-statistics evaluate the difference in gene frequencies between the populations (Weir and Cockerham, 1984); while in the ϕ -statistics, molecular distances among microsatellite alleles are also taken into consideration, which are calculated as the square of the difference in repeat number (Michalakis and Excoffier, 1996). The significance of the variance components was estimated by a nonparametric permutation procedure, as implemented in Arlequin 1.0 (Schneider et al., 1997). F_{ST} and ϕ_{ST} values, along with the probabilities, are shown in Table 3. The locus showing the highest between-population molecular variance is DYS388 ($F_{ST} = 0.38$, $\phi_{ST} = 0.58$). For this marker, a very different allelic distribution is observed between the Tibeto-Burman populations and the rest of the populations

included in the analysis. The 123 bp allele is the most common in the four northeast Indian tribal populations, while it appears in much lower frequency in some of the mainland southeast Asian populations (northern Thai, So, Cambodians, and Chinese). This allele is not observed in most insular southeast Asian populations, or in Japanese, Samoan, and New Guinea Highlanders. The same pattern is observed in DYS395 ($F_{ST} = 0.26$, $\phi_{ST} = 0.27$), at which the 119 bp allele is the most frequent allele found in the Tibeto-Burmans and southern Chinese; while the 123 bp and 127 bp alleles are the common alleles in the remaining populations. On the contrary, DYS391 shows a more homogeneous distribution ($F_{ST} = 0.09$, $\phi_{ST} = 0.12$), where the 283 bp allele is the most prevalent one in almost all the populations, with the exception of Apatani and Cambodians.

The neighbor-joining tree based on the D_A genetic distance is presented in Figure 2. The main division of the tree is observed between mainland populations (with the exception of Northeastern Thai) and a heterogeneous group comprising of the Austronesian populations (Batak, Javanese, Malay, Atayal, and Samoan), Orang-Asli, northeastern Thai, New Guinea Highlanders, and Japanese. This group and the Tibeto-Burman cluster show high bootstrap values (88% and 87%), but the internal bootstrap values are substantially lower, indicating a lack of structure in the tree. The neighbor-joining tree based on the D_{SW} distances (not shown) also indicates a clear separation of the Tibeto-Burman populations (bootstrap value 92%), which cluster with the southern Chinese, as it is observed in the D_A tree.

Another way of depicting the genetic relationships of the populations, without some of the limitations inherent in tree construction, is through the ordination methods. These methods have been designed to interpret multidimensional information, in this case in the form of frequency or distance data, by transforming the original variables into fewer variables with a minimum loss of information. Figure 3 presents a principal coordinate analysis based on the D_A genetic distances. Very similar results were obtained using the D_{SW} genetic distances, or

TABLE 2. Allele frequencies of five Y-chromosome microsatellites in Asian populations

Population	DYS338		DYS390		DYS391		DYS394		DYS395	
	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>
Kachari	120	0.00	187	0.00	275	0.00	243	0.00	115	0.11
	123	0.67	195	0.00	279	0.00	247	0.59	119	0.56
	126	0.00	199	0.00	283	0.85	251	0.26	123	0.15
	129	0.33	203	0.00	287	0.11	255	0.11	127	0.19
	132	0.00	207	0.07	291	0.04	259	0.04	131	0.00
	135	0.00	211	0.63					135	0.00
	138	0.00	215	0.07						
			219	0.22						
			223	0.00						
			227	0.00						
N		27		27		27		27		27
D (Avg = 0.51)		0.46		0.56		0.27		0.59		0.65
Nishi	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.95	195	0.00	279	0.00	247	0.95	119	0.95
	126	0.00	199	0.00	283	0.95	251	0.00	123	0.05
	129	0.05	203	0.00	287	0.05	255	0.05	127	0.00
	132	0.00	207	0.00	291	0.00	259	0.00	131	0.00
	135	0.00	211	0.37					135	0.00
	138	0.00	215	0.58						
			219	0.05						
			223	0.00						
			227	0.00						
N		19		19		19		19		19
D (Avg = 0.20)		0.11		0.56		0.11		0.11		0.11
Apatani	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.79	195	0.00	279	0.00	247	0.20	119	0.93
	126	0.00	199	0.00	283	0.43	251	0.50	123	0.07
	129	0.21	203	0.00	287	0.57	255	0.30	127	0.00
	132	0.00	207	0.14	291	0.00	259	0.00	131	0.00
	135	0.00	211	0.64					135	0.00
	138	0.00	215	0.14						
			219	0.07						
			223	0.00						
			227	0.00						
N		14		14		14		10		14
D (Avg = 0.46)		0.36		0.58		0.53		0.69		0.14
Adi	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.96	195	0.00	279	0.00	247	0.65	119	0.92
	126	0.04	199	0.00	283	0.88	251	0.35	123	0.08
	129	0.00	203	0.00	287	0.12	255	0.00	127	0.00
	132	0.00	207	0.00	291	0.00	259	0.00	131	0.00
	135	0.00	211	0.46					135	0.00
	138	0.00	215	0.54						
			219	0.00						
			223	0.00						
			227	0.00						
N		24		24		24		20		24
D (Avg = 0.29)		0.08		0.52		0.23		0.48		0.16
Japanese	120	0.00	187	0.00	275	0.05	243	0.15	115	0.00
	123	0.00	195	0.00	279	0.05	247	0.05	119	0.25
	126	0.00	199	0.00	283	0.75	251	0.45	123	0.55
	129	0.74	203	0.00	287	0.15	255	0.15	127	0.20
	132	0.26	207	0.32	291	0.00	259	0.20	131	0.00
	135	0.00	211	0.10					135	0.00
	138	0.00	215	0.32						
			219	0.21						
			223	0.00						
			227	0.05						
N		19		19		20		20		20
D (Avg = 0.60)		0.41		0.78		0.43		0.75		0.63
Chinese	120	0.00	187	0.00	275	0.00	243	0.05	115	0.00
	123	0.32	195	0.00	279	0.05	247	0.30	119	0.55
	126	0.00	199	0.00	283	0.80	251	0.45	123	0.20
	129	0.58	203	0.00	287	0.15	255	0.15	127	0.20
	132	0.10	207	0.00	291	0.00	259	0.05	131	0.05
	135	0.00	211	0.45					135	0.00
	138	0.00	215	0.30						
			219	0.25						
			223	0.00						

(Continued)

TABLE 2. (continued)

Population	DYS338		DYS390		DYS391		DYS394		DYS395	
	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>
N		19	227	0.00						
D (Avg = 0.60)		0.58	20	0.68	20	0.35	20	0.72	20	0.65
N Thai	120	0.07	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.25	195	0.00	279	0.00	247	0.36	119	0.32
	126	0.00	199	0.00	283	0.71	251	0.29	123	0.28
	129	0.43	203	0.00	287	0.29	255	0.25	127	0.32
	132	0.07	207	0.04	291	0.00	259	0.11	131	0.04
	135	0.18	211	0.11					135	0.04
	138	0.00	215	0.64						
			219	0.18						
			223	0.04						
			227	0.00						
N		28	28		28		28		28	
D (Avg = 0.64)		0.74	0.56		0.42		0.74		0.74	
NE Thai	120	0.00	187	0.00	275	0.07	243	0.00	115	0.04
	123	0.00	195	0.00	279	0.18	247	0.07	119	0.07
	126	0.00	199	0.00	283	0.54	251	0.68	123	0.25
	129	0.89	203	0.00	287	0.21	255	0.11	127	0.57
	132	0.11	207	0.04	291	0.00	259	0.14	131	0.07
	135	0.00	211	0.28					135	0.00
	138	0.00	215	0.39						
			219	0.25						
			223	0.04						
			227	0.00						
N		28	28		28		28		28	
D (Avg = 0.54)		0.20	0.72		0.65		0.52		0.62	
Cambodian	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.29	195	0.00	279	0.00	247	0.36	119	0.29
	126	0.00	199	0.00	283	0.43	251	0.29	123	0.29
	129	0.64	203	0.00	287	0.57	255	0.36	127	0.36
	132	0.00	207	0.00	291	0.00	259	0.00	131	0.07
	135	0.07	211	0.07					135	0.00
	138	0.00	215	0.71						
			219	0.21						
			223	0.00						
			227	0.00						
N		14	14		14		14		14	
D (Avg = 0.60)		0.54	0.47		0.53		0.71		0.76	
So	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.25	195	0.00	279	0.00	247	0.50	119	0.25
	126	0.00	199	0.00	283	0.63	251	0.37	123	0.25
	129	0.12	203	0.00	287	0.37	255	0.13	127	0.37
	132	0.25	207	0.00	291	0.00	259	0.00	131	0.13
	135	0.38	211	0.43					135	0.00
	138	0.00	215	0.29						
			219	0.29						
			223	0.00						
			227	0.00						
N		8	7		8		8		8	
D (Avg = 0.72)		0.82	0.76		0.54		0.68		0.82	
Orang Asli	120	0.00	187	0.00	275	0.00	243	0.00	115	0.06
	123	0.00	195	0.00	279	0.12	247	0.18	119	0.06
	126	0.12	199	0.00	283	0.88	251	0.76	123	0.41
	129	0.76	203	0.06	287	0.00	255	0.00	127	0.23
	132	0.12	207	0.18	291	0.00	259	0.06	131	0.23
	135	0.00	211	0.06					135	0.00
	138	0.00	215	0.18						
			219	0.53						
			223	0.00						
			227	0.00						
N		17	17		17		17		17	
D (Avg = 0.50)		0.41	0.69		0.22		0.40		0.76	
Malay	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.09	195	0.00	279	0.29	247	0.00	119	0.31
	126	0.00	199	0.00	283	0.57	251	0.57	123	0.69
	129	0.64	203	0.00	287	0.14	255	0.43	127	0.00
	132	0.18	207	0.00	291	0.00	259	0.00	131	0.00
	135	0.00	211	0.31					135	0.00

(Continued)

TABLE 2. (continued)

Population	DYS338		DYS390		DYS391		DYS394		DYS395	
	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>	Size	<i>p</i>
N D (Avg = 0.58) Javanese	138	0.09	215	0.38						
			219	0.31						
			223	0.00						
			227	0.00						
		11	13		14		14		13	
		0.60		0.72		0.62		0.53		0.46
	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.00	195	0.00	279	0.10	247	0.19	119	0.20
	126	0.00	199	0.00	283	0.80	251	0.56	123	0.55
	129	0.59	203	0.00	287	0.10	255	0.25	127	0.10
	132	0.35	207	0.12	291	0.00	259	0.00	131	0.15
	135	0.06	211	0.29					135	0.00
	138	0.00	215	0.35						
N D (Avg = 0.59) Batak			219	0.24						
			223	0.00						
			227	0.00						
		17	17		20		16		20	
		0.56		0.76		0.36		0.63		0.66
	120	0.00	187	0.00	275	0.00	243	0.00	115	0.06
	123	0.00	195	0.00	279	0.33	247	0.11	119	0.22
	126	0.00	199	0.00	283	0.44	251	0.44	123	0.33
	129	0.94	203	0.23	287	0.22	255	0.39	127	0.33
	132	0.06	207	0.18	291	0.00	259	0.06	131	0.06
	135	0.00	211	0.18					135	0.00
	138	0.00	215	0.18						
			219	0.23						
N D (Avg = 0.61) Atayal			223	0.00						
			227	0.00						
		18	17		18		18		18	
		0.11		0.85		0.68		0.67		0.76
	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.00	195	0.00	279	0.00	247	0.00	119	0.00
	126	0.00	199	0.00	283	0.52	251	0.74	123	0.91
	129	0.91	203	0.00	287	0.35	255	0.22	127	0.04
	132	0.09	207	0.00	291	0.13	259	0.04	131	0.00
	135	0.00	211	0.78					135	0.04
	138	0.00	215	0.17						
			219	0.04						
			223	0.00						
N D (Avg = 0.35) Samoan			227	0.00						
		23	23		23		23		23	
		0.17		0.37		0.62		0.42		0.17
	120	0.00	187	0.05	275	0.00	243	0.00	115	0.00
	123	0.00	195	0.10	279	0.05	247	0.10	119	0.00
	126	0.00	199	0.35	283	0.85	251	0.40	123	0.45
	129	0.10	203	0.05	287	0.10	255	0.50	127	0.55
	132	0.35	207	0.05	291	0.00	259	0.00	131	0.00
	135	0.05	211	0.00					135	0.00
	138	0.50	215	0.35						
			219	0.05						
			223	0.00						
			227	0.00						
N D (Avg = 0.57) Highlander		20	20		20		20		20	
		0.65		0.77		0.28		0.61		0.52
	120	0.00	187	0.00	275	0.00	243	0.00	115	0.00
	123	0.00	195	0.00	279	0.00	247	0.18	119	0.05
	126	0.00	199	0.00	283	0.79	251	0.76	123	0.70
	129	0.50	203	0.00	287	0.21	255	0.06	127	0.25
	132	0.39	207	0.00	291	0.00	259	0.00	131	0.00
	135	0.11	211	0.44					135	0.00
	138	0.00	215	0.22						
			219	0.06						
			223	0.28						
			227	0.00						
		18	18		19		17		20	
N D (Avg = 0.51) Global D		0.62		0.71		0.35		0.40		0.47
		0.44		0.65		0.42		0.57		0.53

Note: Size represents the fragment size in base pairs. *p* represents frequency of alleles; the most frequent allele is marked in **boldface**. D is gene diversity, calculated as $D = (1 - \sum p^2) / n - 1$. N is sample size.

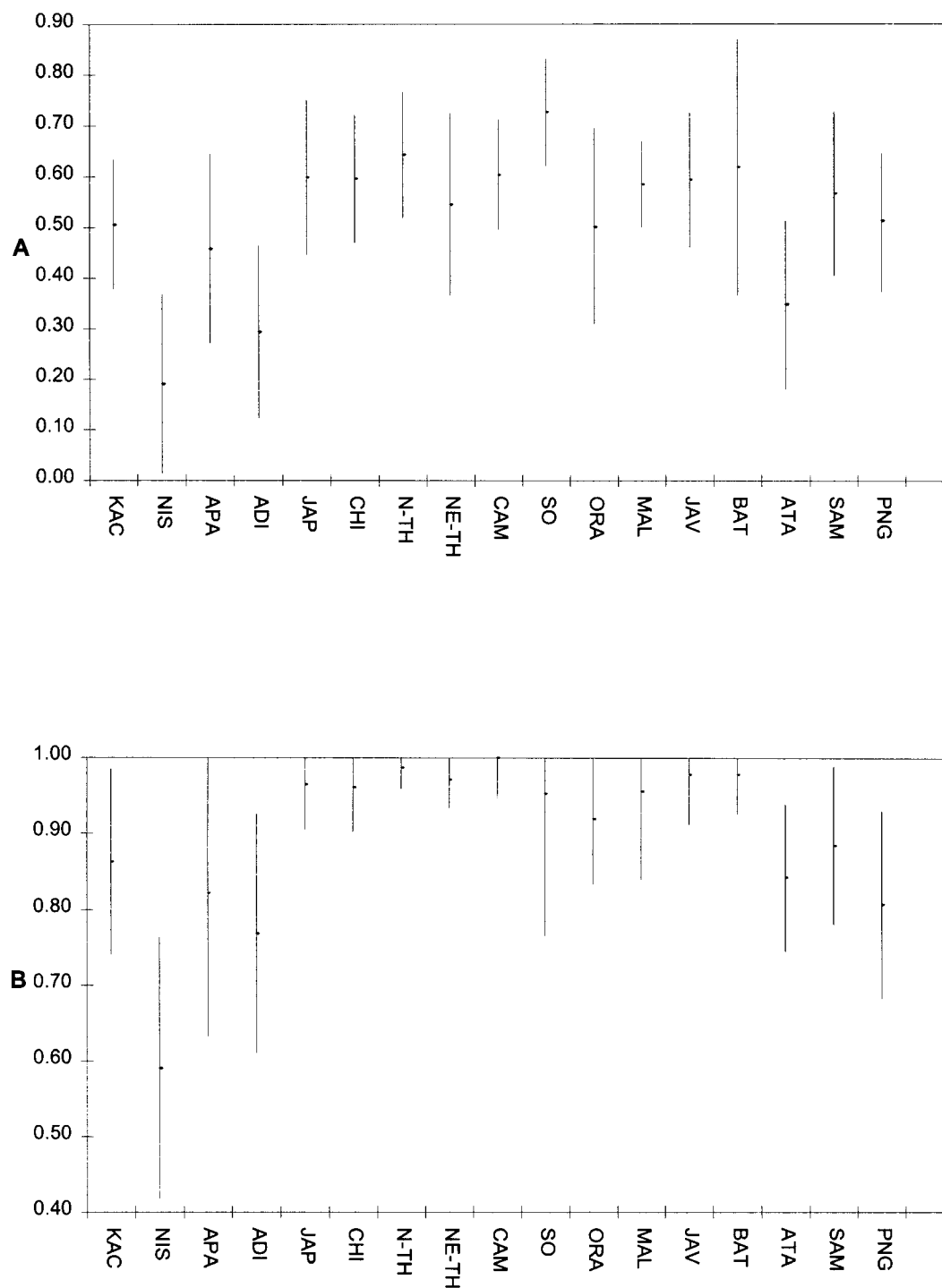


Fig. 1. Gene diversities, including 95% confidence intervals (in the range from 0 to 1), observed in 17 Asian and Pacific populations. **(A)** Based on five Y-specific microsatellite loci allele frequencies; **(B)** based on haplotype frequencies. Note: In figures and appendix, population abbreviations are used as follows: ADI (Adi), APA

(Apatani), ATA (Atayal), BAT (Batak), CAM (Cambodian), CHI (southern Chinese), JAV (Javanese), JAP (Japanese), KAC (Kachari), MAL (Malay), N-TH (northern Thai), NE-TH (northeastern Thai), NIS (Nishi), ORA (Orang Asli), PNG (Papua New Guinea Highlander), SAM (Samoan), SO (So).

TABLE 3. F_{ST} and ϕ_{ST} values for five Y-specific microsatellite loci in 17 Asian and Oceanic populations

Marker	F_{ST}	ϕ_{ST}
DYS388	0.38, $p < 0.001$	0.58, $p < 0.001$
DYS390	0.13, $p < 0.001$	0.22, $p < 0.001$
DYS391	0.09, $p < 0.001$	0.12, $p < 0.001$
DYS394	0.17, $p < 0.001$	0.16, $p < 0.001$
DYS395	0.26, $p < 0.001$	0.27, $p < 0.001$
Average	0.21, $p < 0.001$	0.33, $p < 0.001$
Haplotypes	0.09, $p < 0.001$	0.36, $p < 0.001$

other multivariate methods, such as the correspondence analysis (data not shown). The first three principal coordinates explain more than 64% of the variance (41.9%, 12.3%, and 10.4%). The Tibeto-Burman speaking Adi and Nishi occupy extreme negative values on the first axis. The other two Tibeto-Burman populations, Kachari and Apatani, also in negative positions on the first PC, are closer to a group of mainland southeast Asian populations, including southern Chinese, northern Thai, Cambodian, and So, the latter three populations located close together in the representation. A third group includes Japanese, Papua New Guinea Highlanders, Batak, Javanese, Malays, and northeastern Thai, with two tribal populations (Orang-Asli and Atayal) in opposite locations on the third principal coordinate. Finally, the Samoan population is an outlier, distinct from the other populations of insular southeast Asia. In fact, the Samoans present unique characteristics in their allele distributions for some of the markers. At the DYS388 locus, the 138 bp allele is the most common allele in Samoans ($p = 0.50$), while it has not been observed in any of the other Asian populations, with the exception of the Malays. Similarly at DYS390, the 199 bp allele is observed only among the Samoans. They also show clear differences from the other populations at DYS394, with very high frequencies for the 255 bp allele.

Haplotype analysis

The haploid nature of the non-recombining portion of the Y chromosome permits generation of discrete haplotypes. We have detected 123 different haplotypes in our sample of 314 chromosomes. Eighty-seven haplotypes were seen only in one population. Of the 36 remaining haplotypes, 18

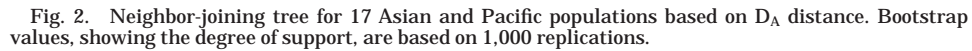
were observed in two populations and 18 in more than two populations. It should, however, be noted that some low-frequency haplotypes probably remained undetected due to small sample sizes of the examined populations. Nevertheless, the distribution of some of the observed haplotypes seems to be restricted to specific linguistic or geographic groups, but in other cases they show a wide geographical distribution in southeast Asia and the Pacific (see Appendix). The gene diversity analysis shows the same trends previously described in the analysis of the individual loci (see Figure 1B). Most mainland populations and some insular populations show diversities higher than 0.95. Among the Tibeto-Burman populations, Kachari and Apatani show slightly lower gene diversities (0.86 and 0.82, respectively), but Nishi and Adi present very low diversities (0.59 and 0.77, respectively), with significant p values in several of the comparisons with other populations. The gene diversities observed in PNG Highlanders, Atayal and Samoans are also lower than 0.90.

The analysis of molecular variance also corroborates the results presented in the previous section, indicating a substantial genetic differentiation of the Y-specific microsatellite haplotypes in this geographical area ($F_{ST} = 0.09$, $p < 0.001$ and $\phi_{ST} = 0.36$, $p < 0.001$).

Finally, both the D_A trees constructed using the haplotype frequencies and the principal coordinate analysis (data not shown) show a basic conformity with the results obtained treating the loci individually. The mainland populations (with the exception of the northeastern Thai) form two clusters, one including the Tibeto-Burman populations, and the other southern Chinese, Cambodian, northern Thai, and So. The remaining populations cluster together in a heterogeneous group.

DISCUSSION

We have characterized genetic variation at five microsatellite loci located in the non-recombining portion of the Y chromosome (DYS388, DYS390, DYS391, DYS394, and DYS395) in 17 Asian and Pacific populations drawn from a broad geographical area and belonging to different linguistic families.



From the analysis of the allele frequency distributions and gene diversities in these populations, it can be inferred that some of the population groups have remained isolated with smaller male effective sizes, or have suffered bottlenecks in the past. This seems to be the case in two tribal Tibeto-Burman populations from northeast India (Nishi and Adi), which show very low gene diversities, and to a lesser extent in the Taiwanese aborigines (Atayal). Other tribal or insular populations (Orang-Asli, Highlanders, and Samoans) also show slightly lower gene diversities than the rest of the populations. The average values of the F_{ST} and ϕ_{ST} statistics indicate a substantial genetic differentiation among the studied

Both the neighbor-joining trees and the principal coordinates representation based on the Y-specific microsatellite loci show a separation of the mainland and insular populations. Among the mainland populations,

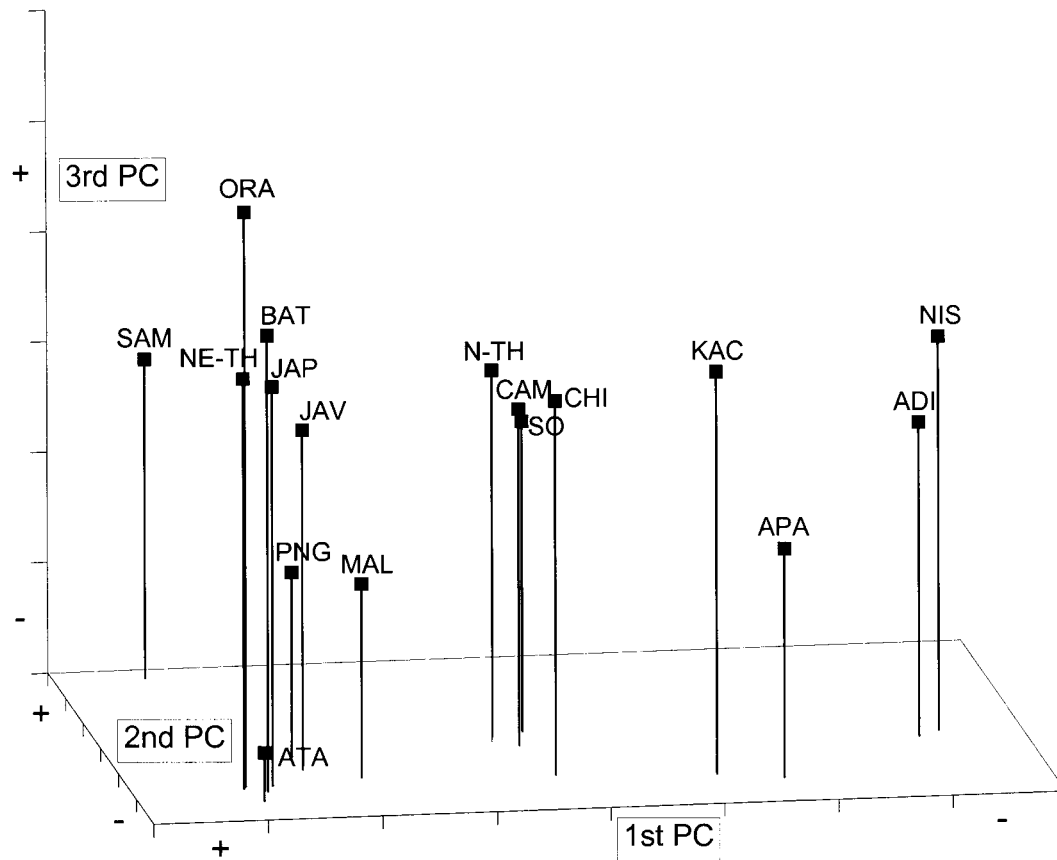


Fig. 3. Principal coordinates representation of 17 Asian and Pacific populations.

the Tibeto-Burman Indian tribal populations (Kachari, Apatani, Nishi, and Adi) cluster together, with Nishi and Adi showing very long branches in the NJ trees and occupying marginal locations in the PC representation, most likely due to the effects of a stronger genetic drift. The southern Chinese cluster with the Tibeto-Burman populations in the NJ tree, but both the principal coordinates plot and the D_A tree based on the haplotype frequencies show that they are located close to three additional mainland southeast Asian populations: northern Thai (Daic speakers), and Cambodian, and So (Austroasiatic speakers). It is noteworthy that the second Daic-speaking population, northeastern Thai, does not cluster with the other mainland populations. Similarly, the Malayan aborigines (Orang Asli or Semai), who speak an Austroasiatic language and

live in isolated pockets in the interior of the Malay Peninsula are quite different from the other two Austroasiatic-speaking populations. The third group is quite heterogeneous, which include, in addition to northeastern Thai and Orang Asli, the Austronesian speakers (Malay, Javanese, Batak and Atayal from Taiwan), PNG Highlanders, and Japanese. Finally, the Samoans appear as an outlier, a reflection of their unique features with regard to the gene frequencies, possibly resulting from founder effects during the process of the colonization of the Pacific by the Austronesian populations.

The patterns of population relationships emerging from the analysis of Y-specific microsatellites have not always been as expected. On one hand, we have observed a clear affinity among the mainland popula-

tions, especially the northeast Indian groups, who are geographically as well as ethno-historically related. On the other, there is a cluster that includes populations known to be of diverse origins. For example, allelic distributions at these loci in both the PNG Highlanders and Japanese are not very different from those observed in the Austronesian populations, all of whom have clustered together in the dendrograms. However, known ethno-history as well as previous studies of these two populations based on a variety of other genetic markers do not indicate the observed relationships (Cavalli-Sforza et. al., 1994; Nei, 1995; Roberts-Thomson et. al., 1996; Omoto and Saitou, 1997; Hammer et. al., 1997). It is obvious that the use of hypervariable microsatellites alone, with their high mutation rate and the convergent nature of their mutational processes could obscure the true picture of

genetic relationships and affinities among populations. Incorporation of additional, more stable, Y-specific polymorphism would be necessary to clarify the complex migratory histories that have characterized southeast Asia and the south Pacific.

ACKNOWLEDGMENTS

We thank Drs. P.E. Smouse, J.C. Long, S-J Yin, and T. Miki for providing the New Guinea Highlander, Chinese, Atayal, Chinese and the Japanese samples. We thank Ling Mei Yu and Elisa Mueller Heidreich for technical assistance. We appreciate the assistance and hospitality provided by the authorities of the Ramakrishna Mission Centers in Arunachal Pradesh to NS in the collection of samples. This work was supported by grants from the National Science Foundation SBR-9600910 and from the National Institutes of Health GM-45861 to RD.

Appendix 1. Observed haplotypes in Asian and Pacific populations

Locus names and alleles (in base pairs)					Populations																		
DYS388	DYS390	DYS391	DYS394	DYS395	KAC	NIS	APA	ADI	JAP	CHI	N-TH	NE-TH	CAM	SO	ORA	MAL	JAV	BAT	ATA	SAM	PNG	Total	
120	215	287	259	127							1											1	
120	219	283	251	123							1											1	
123	211	283	247	115																		3	
123	211	283	247	119	10	6		4		3	1			1								25	
123	211	283	247	123	1																	1	
123	211	283	251	119	1		1	2														4	
123	211	283	251	123				2														2	
123	211	287	247	119	1	1																2	
123	211	287	251	119			4															4	
123	215	283	247	119	1	11	2	9			2											25	
123	215	283	251	119	1			1		2	1			1		1						7	
123	215	283	255	123										1								1	
123	215	287	247	119							1			1								2	
123	215	287	251	119				1														1	
123	219	283	247	119						1	2				1							4	
123	219	287	247	119										1		2						1	
126	207	283	247	123																		2	
126	211	283	251	119				1								2						1	
129	187	283	247	123																1		1	
129	203	279	255	127														2				2	
129	203	283	251	131														1				1	
129	207	279	259	119							1											1	
129	207	283	243	127					1													1	
129	207	283	247	115											1			1				2	
129	207	283	247	127					1									1				2	
129	207	283	251	119	1																	1	
129	207	283	251	123					3													3	
129	207	283	251	127	1												1					2	
129	207	283	255	119			2											1				3	
129	207	283	255	131							1											1	
129	207	287	251	123					1													1	
129	211	279	247	115								1										1	
129	211	279	247	123								1										1	
129	211	279	251	119														2				2	
129	211	279	255	119																		1	
129	211	283	243	119						1							1					1	
129	211	283	251	123					1			2			1	1	1		6		1	13	
129	211	283	251	127							1											1	
129	211	283	255	119										1			1					2	
129	211	283	255	123					1			1				1			2			5	
129	211	283	259	119							1											1	
129	211	283	259	127								1										1	
129	211	287	247	123						2												2	
129	211	287	251	119									1									1	
129	211	287	251	123														1	7			8	

(Continued)

Appendix 1. (continued)

Locus names and alleles (in base pairs)					Populations																		Total
DYS388	DYS390	DYS391	DYS394	DYS395	KAC	NIS	APA	ADI	JAP	CHI	N-TH	NE-TH	CAM	SO	ORA	MAL	JAV	BAT	ATA	SAM	PNG		
129	211	287	259	123	1																		1
129	211	291	251	123																3			3
129	215	275	251	123								1											1
129	215	279	251	123								1											1
129	215	279	255	119													1						1
129	215	279	259	119												1		1					2
129	215	283	247	123													1				1		2
129	215	283	247	127										1									1
129	215	283	251	119					1		1												2
129	215	283	251	123							1							1					2
129	215	283	251	127						1	1	4	1		1		1		1				10
129	215	283	255	123																			2
129	215	283	255	127						1	1	1						1	1				3
129	215	283	259	119						1													1
129	215	287	247	127																			1
129	215	287	251	127								2											2
129	215	287	255	123							1		1										2
129	215	287	255	127							1		1										2
129	215	287	255	131									1										1
129	215	287	259	127							1												1
129	219	279	251	119						1													1
129	219	279	255	123												2							2
129	219	279	259	123					1														1
129	219	283	251	123	1									2			2						5
129	219	283	251	127	1						1	3			3								8
129	219	283	251	131											4		2						6
129	219	283	255	123	1	1																	2
129	219	283	255	127	1							1											2
129	219	283	259	123					3														3
129	219	283	259	127								1											1
129	219	283	251	119						2													2
129	219	287	247	123													1						1
129	219	287	251	123						1				1		1							3
129	219	287	251	127									1	1	1		1						4
129	219	287	255	123			1																2
129	219	287	255	127	1						1							1		1			3
129	219	287	259	127								1											1
129	219	291	251	127	1																		1
129	223	279	251	127								1											1
129	223	283	247	119																		1	1
129	223	287	247	123																		2	2
129	223	287	251	123																		2	2
129	227	283	251	119					1														1
132	195	283	251	127																1			1
132	203	279	251	123											1								1

(Continued)

[illegible]

LITERATURE CITED

- Bellwood P. 1978. Man's conquest of the Pacific: the prehistory of southeast Asia and Oceania. New York: Oxford University Press.
- Bellwood P. 1989. The colonization of the Pacific: some current hypotheses. In: Hill AVS and SW Serjeantson, editors. The colonization of the Pacific: a genetic trail. Oxford: Clarendon. p 1–59.
- Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 19:233–257.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton: Princeton University Press.
- Chen LZ, Eastaale S, Board PG, Kirk RL. 1992. Genetic affinities of oceanic populations based on RFLP and haplotype analysis of genetic loci on three chromosomes. *Hum Biol* 64:1–15.
- Deka R, Jin L, Shriver MD, Yu LM, Saha N, Barrantes R, Chakraborty R, Ferrell RE. 1996. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res* 6:1177–1184.
- Dorit R, Akashi H, Gilbert W. 1995. Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* 268:1183.
- Hammer MF. 1995. A recent common ancestry for human Y chromosomes. *Nature (London)* 378:376–378.
- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, SL Zegura SL. 1997. The geographic distribution of human Y chromosome variation. *Genetics* 145:787–805.
- Jakubiczka S, Arnemann J, Cooke HJ, Krawczak M, Schmidtke J. 1989. A search for restriction fragment length polymorphism on the human Y chromosome. *Hum Genet* 84:86–88.
- Jobling M, Tyler-Smith C. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449–456.
- Lum JK, Cann R. 1998. mtDNA and language support a common origin of Micronesian and Polynesians in island southeast Asia. *Am J Phys Anthropol* 105:109–119.
- Malaspina P, Persichetti F, Novelletto A, Iodice C, Terrenato L, Wolfe J, Ferraro M, Prantero G. 1990. The human Y chromosome shows a low level of DNA polymorphism. *Ann Hum Genet* 54:297–305.
- Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* 57:403–414.
- Michalakis Y, Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142:1061–1064.
- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Nei M. 1995. The origin of human populations: genetic, linguistic and archeological data. In: Brenner S Hanhara K. editors. The origin and past of modern humans as view from DNA. Singapore: World Scientific. p 71–91.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* 19:153–170.
- Omoto K, Saitou N. 1997. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am J Phys Anthropol* 102:437–446.
- Roberts-Thomson JM, Martinson JJ, Norwich JT, Harding RM, Clegg JB, Boettcher B. 1996. An ancient common origin of aboriginal Australians and New Guinea Highlanders is supported by alpha-globin haplotype analysis. *Am J Hum Genet* 58:1017–1024.
- Roewer L, Kayser M, Deltjes P, Nagy M, Bakker E, Krawczak M, Knijff de P. 1996. Analysis of molecular variance (AMOVA) of Y-chromosome specific microsatellites in two closely related human populations. *Hum Mol Genet* 5:1029–1033.
- Rohlf FJ. 1992. NTSYS-pc version 1.70.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol. Evol.* 4:406–425.
- Santos FR, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen JT, Pena SDJ. 1996. Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. *Hum Genet* 97:309–313.
- Schneider S, Kueffer J-M, Roessli D, Excoffier L. 1997. Arlequin ver 1.0.
- Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 12:914–920.
- Spurdle AB, Jenkins T. 1992. The search for Y chromosome polymorphism is extended to negroids. *Hum Mol Genet* 1:169–170.
- Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M. 1995. The origins of Polynesians: an interpretation from mitochondrial lineage analysis. *Am J Hum Genet* 57:1463–1475.
- Takezaki N, Nei M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.
- Underhill PA, Jin L, Zemmans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implication for human evolutionary history. *Proc Natl Acad Sci USA* 93:196–200.
- Voegelin CF, Voegelin FM. 1977. Classification and index of the world's languages. New York: Elsevier.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Whitfield LS, Sulston JE, Goodfellow PN. 1995. Sequence variation of the human Y chromosome. *Nature (London)* 378:379–380.